

Release 3.1A John F. Collins, Blocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPerch_un n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sun Mar 7 02:33:19 1999; MasPar time 389.26 Seconds
1390.239 Million cell updates/sec
Tabular output not generated.

Title: >US-08-962-094-1
Description: (1-229) from US08962094.seq
Perfect Score: 223
N.A. Sequence: 1 CGGCTCGAGCTCTTAGGCTT.....TCGNACCACTGGGACCACTG 229
Comp: GCCGAGCTCGAGATCCGAA.....ACGNTGTGACGCTGTGTGAC

Scoring table: TABLE default
Gap 6

* Match STD : Dbase 0; Query 0

Searched: 602357 segs, 1181590623 bases x 2

Post-Processing: Minimum Match 50%
Listing first 1000 summaries

Database: emb157
1:em_ba 2:em_fun 3:em_htg 4:em_hun1 5:em_hun2 6:em_in
7:em_om 8:em_ov 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_un 15:em_vl
Database: genbank110
16:gb_pat 17:gb_ba2 18:gb_htg 19:gb_in 20:gb_om 21:gb_ov
22:gb_pat 23:gb_ph 24:gb_pl 25:gb_pl 26:gb_pl
27:gb_pr2 28:gb_pr3 29:gb_ro 30:gb_st 31:gb_sts 32:gb_sy
33:gb_un 34:gb_vl

Statistics: Mean 9.529; Variance 5.152; scale 1.850

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description	Pred. No.
No.								

No matches found.

Search completed: Sun Mar 7 02:44:29 1999
Job time : 670 secs.

Arthur L.
08/062094 Page 1
Seq IDs
1-5 & 17-20

WIPSETH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 02:54:43 1999; Maspar time 43.34 Seconds

718.825 Million cell updates/sec
Tabular output not generated.

Title: >US-08-962-094-1
Description: (1-229) from US08962094.seq
Perfect Score: 223
N.A. Sequence: 1 CGGCTCGAGCTCTAGGCTT.....TGCNACGACTGCAGCACTG 229
Comp: GCCGAGCTCGAATCCGA.....ACGNTGGTGAAGCTGGTGAC

Scoring table: TABLE default
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 188442 segs, 6802649 bases x 2

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 7.735; Variance 5.150; scale 1.502

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description	Pred. No.
------------	-------------	-------	--------	----	-------------	-----------

No matches found.

Search completed: Sun Mar 7 02:58:23 1999
Job time : 220 secs.

(7M)

	RESULT	4		EST	
LOCUS	W72838	432 bp	mRNA		16-OCT-1996
DEFINITION	z657g12.s1 Soares fetal heart NDHH19W Homo sapiens cDNA clone 344806 3' similar to contains element MER40 repetitive element ; mRNA sequence.				
ACCESSION	W72838				
NID	G1382814				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 432)				
AUTHORS	Hillier,L., Clark,N., Dubucq,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kuecaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.				
TITLE	The WashU-Merck EST Project				
JOURNAL	Unpublished (1995)				
COMMENT	<u>Contestant notice by</u>				

```

FEATURES
  source
    WashU-Merck EST Project
    Washington University School of Medicine
    4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
    Tel: 314 286 1800
    Fax: 314 286 1810
    Email: est@watson.wustl.edu
    This clone is available royalty-free through LNL; contact the
    IMAGE Consortium (info@image.lnl.gov) for further information.
    Insert Length: 508
    Seq primer: mob.REGA+ET
    High quality sequence stop: 394.
    location/Qualifiers
      1..432

```

```

/organ="heart; Vector: pF773D (Pharmacia) with a
/nots="; modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'-
TGTTTACCACTCTGAAGTGTGGAGCGCGCCGACATCTTTTCTTTTCTTTT 3']
double-stranded cDNA was site selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pTV73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M. Fatima Ronaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."
/db_xref="taxon:9606"
/clone="344806"
/clone_lib="Soares fetal heart NBHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
complement(cl_1..>432)

```

Query Match	75.8%;	Score 169;	DB 21;	Length 432;
Best Local Similarity	96.7%;	Pred. No. 4.77e-260;		
Matches 175;	Conservative	0;	Mismatches 5;	Indels 1;
				Gaps 1;
Db	252	CAGTGCATCGACATGGCTTGACAGACAGTCAGTTCACGATCAGGCGCTTCATCATTACGACAG	311	

Cp 229 CATGGTCCCACTGGTTCGACAGCAGCAGTGGTTCACACNTCAGGGGCTTCATCATCAGCAG 170
 |||||
 Db 312 GACCAGTAACTGGATACGTCTCACTGGAGACAGCTGTGTGGGATTCGGGGCAGAGACC 371
 |||||
 Cp 169 GACCACTA-CGTGGATNCGTGTCAAGCTGNNAGCAGCTGTGTGCGAATCTCGGGCAGAGACC 111
 |||||
 Db 372 AGAAGATGGAACTCCCAAGATGCCAGAGACTGCTAAAGAACTCATATGGTGGTGACATGA 431
 |||||
 Cp 110 AGANNAGTGGAACTCCCAAGATGCCAGAGACTGCTAAAGAACTCATATGGTGGTGACATGA 51
 |||||
 Db 432 A 432
 Cp 50 A 50

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	
5	AI127172	443 bp mRNA	EST	AI127172	q9b9a08.x1 Soares fetal heart_NDHH19W Homo sapiens cDNA clone IMAGE:1708214 3', mRNA sequence.	AI127172	93595686	EST.	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	1 (bases 1 to 443) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . NCI-CGAP Cancer Institute, Cancer Genome Anatomy Project (CGAP). Tumor Gene Index Unpublished (1997)	AI127172

~~Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@hni.gov
This clone is available royalty-free through LINT; contact the
IMAGE Consortium (info@image.lni.gov) for further information.
Sed primer: -40m13 fwd, RT from Amer sham.
Location/Qualifiers
1. 443~~

```

/organism="Homo sapiens"
/ncbi-organ="heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site1: Not I; Site2: Eco RI; 1st
strand cDNA was primed with a Not I - Oligo(dT) primer 1'
TGTTCACATCTGCAGCTGGAGGCGCGGACATCTTTTCTTTTCTT 3'),
double-stranded cDNA was size selected, ligated to Eco R
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library was constructed by
M.Patina Bonaldi. This library was constructed from the
same fawn as the fetal lung library. Soares fetal lung
NBH119W."
/db_xref="taxon:9606"
/clone="IMAGE:1708214"
/clone_1lb="Soares_fetal_heart_NBH119W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
129 a 80 c 129 g 103 t 2 others

```

	Query Match	71.3%;	Score 159;	DB 17;	Length 443;
	Best Local Similarity	95.3%;	Pred. No. 3.48e-241;		
	Matches	162;	Conservative	0;	Mismatches 8; Indels 0; Gaps 0;
Db	274	CAGTGTGCAGAGCGATTGCAGCAGCAGTGTGTTTCAGCATCAGGGGCTTCATCATCAGCAG	333		
CP	229	CAGTGTGCAGAGCGTTCAGCAGCAGCAGCAGTGTGTTTCAGTTCAGGGGCTTCATCATCAGCAG	170		
Db	334	GACCAATGCTGATACGTGTAGCTGAGACAGAGTGTTCGGATTCTCGGACAGACCA	393		

Search completed: Sun Mar 7 02:54:19 1999
Job time : 566 secs.

CP 169 GACCACTAGCTGATNGCTGTCAGCTGNACAGCTGTTGCGATTCTGGCAGAGACCA 110
Db 394 GAAAGATGGAGCTGCCAGAGACTACAGAGCTGCTAAGAACTTAATGCTG 443
CP 109 GANAGATGGAATCCCAAGAGTACCAGAGCTGCTAAGAACTTCATGCTG 60

RESULT 6
LOCUS A1139456 402 bp mRNA EST 23-SEP-1998
DEFINITION qc20901.x1 Soares_fetal_heart_NBHL19W Homo sapiens CDNA clone
IMAGE:1710192 3 similar to contains element MER40 repetitive
element ;, mRNA sequence.
A1139456
ACCESSION 93645428
NID
KEYWORDS
SOURCE
ORGANISM

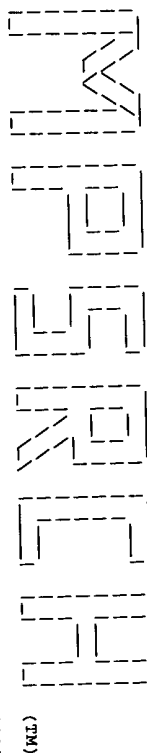
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 402)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strauberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strauberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (infoimage.lnl.gov) for further information.
Seq primer: -40m13 fwd. B1 from Amersham
High quality sequence stop: 398.
Location/Qualifiers
1. 402

FEATURES

Source
/organism="Homo sapiens"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand CDNA was primed with a Not I - Oligo(dT) primer (5'
TCTTACCACTCTGAAGTGGAGCGCGCCGATCTTTTCTTTTCTTTT 3'),
double-stranded CDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Col - 5. Library constructed by
Y. Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."
/db_xref="taxon:9606"
/clone="IMAGE:1710192"
/clone_1lb="Soares_fetal_heart_NBHL19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
BASE COUNT 115 a 72 c 115 g 100 t
ORIGIN

Query Match 63 2%; Score 141; DB 18; Length 402;
Best Local Similarity 95.9%; Pred. No. 2.00e-207;
Matches 142; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Db 255 CAGTGGTCGAGTGGTTCACAGCAGTGTTCAGATCAGGGGCTTCATTATCAGCAG 314
CP 229 CAGTGGTCGAGTGGTTCACAGCAGTGTTCAGATCAGGGGCTTCATTATCAGCAG 170
Db 315 GACCACTAGCTGATGCTGACGTGAGCAGCTGTTGCGATTCTGGCAGAGACCA 374
CP 169 GACCACTAGCTGATGCTGACGTGAGCAGCTGTTGCGATTCTGGCAGAGACCA 110
Db 375 GATAGATGAAGTCCCAAGAGTACCA 402
CP 109 GANAGATGGAATCCCAAGAGTACCA 82



 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mparc_un n.a. - n.a. database search, using Smith-Waterman algorithm
 Run on: Sun Mar 7 02:58:46 1999; Maspar time 514.57 Seconds
 1414.510 Million cell updates/sec
 Tabular output not generated.

Title: >US-08-962-094-2
 Description: (1-308) from US08962094.seq
 Perfect Score: 308
 N.A. Sequence: 1 TAGGCTTGAAGCATTG.....TTGGGATCTCCGATGCT 308
 Comp: ATCCGAACCTGCTAATAAC.....AACCCCTAGAAAGCTTACCA

Scoring table: TABLE default
 Gap 6

Nmatch STD: Dbase 0; Query 0

Searched: 602357 seqs, 1181590623 bases x 2

Post-processing: Minimum Match 50%
 Listing first 1000 summaries

Database: emb157
 Item: ba 2:em_fun 3:em_htg 4:em_hun1 5:em_hun2 6:em_in
 7:em_om 8:em_or 9:em_ov 10:em_pat 11:em_ph 12:em_pl
 13:em_ro 14:em_un 15:em_v1
 genbank110
 16:gb_ba1 17:gb_ba2 18:gb_htg 19:gb_in 20:gb_om 21:gb_ov
 22:gb_pat 23:gb_ph 24:gb_pl 25:gb_pl2 26:gb_pl1
 27:gb_pr2 28:gb_pr3 29:gb_ro 30:gb_st 31:gb_sts 32:gb_sy
 33:gb_un 34:gb_v1

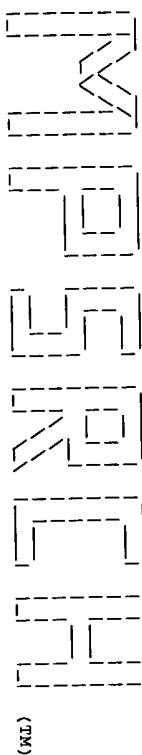
Statistics: Mean 9.810; Variance 5.472; scale 1.793
 Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description	Pred. No.

No matches found.
 Search completed: Sun Mar 7 03:11:01 1999
 Job time : 735 secs.

12



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpserch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 03:23:57 1999; Maspar time 58.14 Seconds
720.802 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-2
Description: (1-308) from US08962094.seq
Perfect Score: 308 1 TAGGCTTGACGATTTTG.....TTGGGATCTTCGATAGT 308
N.A. Sequence: ATCCGAACTGCTTAAC.....AACCCTAGAAGGCTTACCA
Comp:

Scoring table: TABLE default
Gap 6

Mmatch STD: Dbase 0; Query 0

Searched: 188442 segs, 68026449 bases x 2

Post-processing: Minimum Match 508
Listing first 1000 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 7.972; Variance 5.293; scale 1.506

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description	Pred. No.
------------	-------------	-------	--------	----	-------------	-----------

~~No matches found.~~

Search completed: Sun Mar 7 03:27:25 1999
Job time: 208 secs.

Db 181 GCGGCGCAACCACTGCGGACCACTGCGGCTCCACCACTGCACCAACCGCGCTTTACC 240
Oy 192 GCGGCGCAACCACTGCGGACCACTGCGGCTCCACCACTGCACCAACCGCGCTTTACC 251
Db 241 ACGGCGCGTAAGCAATTCCAGTTTACCCAAATGGGTGGGATCTCCCAATGGT 297
Oy 252 ACGGCGCGTAAGCAATTCCAGTTTACCCAAATGGGTGGGATCTCCCAATGGT 308

RESULT	3
LOCUS	AA340069
DEFINITION	EST 294 bp mRNA
ACCESSION	EST45219 Fetal skin Homo sapiens CDNA 5' end similar to similar to
NID	sport 2, submandibular gland protein, mRNA sequence.
KEYWORDS	AA340069
SOURCE	g1992307
ORGANISM	EST.
	human.
	Homo sapiens
	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
	Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;
REFERENCE	1 (bases 1 to 294)
AUTHORS	Adams,M.D., Kerlavave,A.R., Fleischmann,D.B., White,O., Miller,L.,

FEATURES	Location/Qualifiers
source	1. .436

TITLE	Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
JOURNAL	Nature 377 (6547 Suppl.), 3-174 (1995)
MEDLINE	96026280
COMMENT	Other ESTs: THC117518

The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkejavet@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.

FEATURES
SOURCE

```

/organism="Homo sapiens"
/notes="Organ : skin; Vector: plasmidscript SK-; Site:1.
ECORI; Site_2: XhoI"
/db_xref="ATCC (inhost):141700"
/db_xref="taxon:9606"
/clone_1fb="Fetal skin"
/tissue_type="epithelium"
/cell_type="epithelial cell"
/dev_stage="fetus"
<1. .>294
60 a 91 c 58 g 82 t 3 others
BASE COUNT
mRNA
ORIGIN

```

Query Match	94.5%	Score 291;	DB 25;	Length 294;
Best Local Similarity	99.0%;	Pred. No. 0.00e+00;		
Matches 291;	Conservative	0;	Mismatches 3;	Indels 0;
				Gaps 0;

Db 1 CTTTGAAGCATTTTNTCTGTCCTGATCTTCAGGTACACCAACCATGAAGTCTTAGC 60
 CTTTGAAGCATTTTNTCTGTCCTGATCTTCAGGTACACCAACCATGAAGTCTTAGC 64
 5 CTTTGAAGCATTTTNTCTGTCCTGATCTTCAGGTACACCAACCATGAAGTCTTAGC 64
 Db 61 AGTCTGAGTACTCTTGGAGTTTCATCTTTTGTCTCTGGCCAGAAATCCAGAACAGC 120
 AGTCTGAGTACTCTTGGAGTTTCATCTTTTGTCTCTGGCCAGAAATCCAGAACAGC 124
 65 AGTCTGAGTACTCTTGGAGTTTCATCTTTTGTCTCTGGCCAGAAATCCAGAACAGC 124
 Db 121 TGTCTCAGCTGACAGTATCCAGTACTGCTCTGATGATGAAGCCCTGATGCTGA 180
 TGTCTCAGCTGACAGTATCCAGTACTGCTCTGATGATGAAGCCCTGATGCTGA 184
 125 TGTCTCAGCTGACAGTATCCAGTACTGCTCTGATGATGAAGCCCTGATGCTGA 184
 Db 181 AACCACTGCTGCTGCAACCACTGCGACCACTGCTGCTCTACCACTGCAACCACTGCTG 240
 AACCACTGCTGCTGCAACCACTGCGACCACTGCTGCTCTACCACTGCAACCACTGCTG 244
 185 AACCACTGCTGCTGCAACCACTGCGACCACTGCTGCTCTACCACTGCAACCACTGCTG 244
 Db 241 TTTTACCACTGCTGCTGCAACCACTGCTGCTCTACCACTGCAACCACTGCTGCTG 294
 TTTTACCACTGCTGCTGCAACCACTGCTGCTCTACCACTGCAACCACTGCTGCTG 298
 245 TTTTACCACTGCTGCTGCAACCACTGCTGCTCTACCACTGCAACCACTGCTGCTG 298

RESULT 4 W72838 432 bp mRNA EST 16-OCT-1996
 LOCUS Z657912.s1 Soares fetal heart NBH19W Homo sapiens cDNA clone
 DEFINITION 344806 3' similar to contains element MER40 repetitive element ;,
 mRNA sequence.
 W72838
 ACCESSION 91382814
 NID EST.
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Eukaryota; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 432)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, R., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Merritt, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Woldmann, P. and Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 508 Std Error: 0.00
 Seq primer: mob.REGA+ET
 High quality sequence stop: 394.

FEATURES
 source 1..432
 Location/Qualifiers
 1..432
 /organism="Homo sapiens"
 /note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTCACATCTGAGAGGAGGAGGCGGCGCATCTTTTCTTTTCTTTT 3'), double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBH19W."
 /db_xref="taxon:9606"
 /clone="344806"
 /clone_lib="Soares fetal heart NBH19W"

/sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DB10B (ampicillin resistant)"
 complement(<1..>432)
 BASE COUNT 128 a 78 c 124 g 102 t
 ORIGIN

Query Match 85.1%; Score 262; DB 21; Length 432;
 Best Local Similarity 98.9%; Pred. No. 0.00e+00;
 Matches 270; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Db 160 ACCATTCGGGAGATCCCAACCCATTTGGGTAAAGTGAATCTCTTACGACAGTGT 219
 C 308 ACCATTCGGGAGATCCCAACCCATTTGGGTAAAGTGAATCTCTTACGACAGTGT 249
 Db 220 AGAAGCAGCGGTGTTGCTGAGTGAAGCAGAGTGTGCTGAGTGTGCTGAGTGT 279
 C 248 AGAAGCAGCGGTGTTGCTGAGTGAAGCAGAGTGTGCTGAGTGTGCTGAGTGT 189
 Db 280 GGTTCAGCATCAGGGGCTTCATCATCAGCAGACAGTGAAGTGTGATGCTGAGTGT 339
 C 188 GGTTCAGCATCAGGGGCTTCATCATCAGCAGACAGTGAAGTGTGATGCTGAGTGT 130
 Db 340 GAGCAGCTGTGCTGAGTGTGCTGAGCAGACAGTGAAGTGAAGTGAAGTGAAGTGA 399
 C 129 GAGCAGCTGTGCTGAGTGTGCTGAGCAGACAGTGAAGTGAAGTGAAGTGAAGTGA 70
 Db 400 GGAAGCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 432
 C 69 GGAAGCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 37

RESULT 5 A1127172 443 bp mRNA EST 11-SEP-1998
 LOCUS gp99a08.x1 Soares fetal heart NBH19W Homo sapiens cDNA clone
 DEFINITION IMAGE:11708214 3', mRNA sequence.
 A1127172
 ACCESSION 93595686
 NID EST.
 KEYWORDS human.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Eukaryota; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 443)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40m3 fwd. ET from Amersham.

FEATURES
 source 1..443
 Location/Qualifiers
 1..443
 /organism="Homo sapiens"
 /note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5' TGTTCACATCTGAGAGGAGGAGGCGGCGCATCTTTTCTTTTCTTTT 3'), double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBH19W."
 /db_xref="taxon:9606"
 /clone="IMAGE:11708214"
 /clone_lib="Soares fetal heart NBH19W"

```
/sex="unknown"  
/dev_stage="19 weeks"  
/lab_host="DH10B (ampicillin resistant)"  
BASE COUNT      129 a      80 c      129 g      103 t      2 others  
ORIGIN
```

```
Query Match      81.5%; Score 251; DB 17; Length 443;  
Best Local Similarity 98.5%; Pred. No. 0.00e+00;  
Matches 255; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Db 185 ATTGGGAGATCCCAACCCATTGGGTAAACCTGAATGCTTTACGACGAGTGTAGA 244  
|||||  
Cp 305 ATTGGGAGATCCCAACCCATTGGGTAAACCTGAATGCTTTACGACGAGTGTAGA 246  
|||||  
Db 245 AGCAGCGGTGTTGCAAGTGTAGAGACAGACAGTGTGTCAGTGTGTCAGCAGTGTG 304  
|||||  
Cp 245 AGCAGCGGTGTTGCAAGTGTAGAGACAGACAGTGTGTCAGTGTGTCAGCAGTGTG 186  
|||||  
Db 305 TTCAACATCAGGCGCTTATCATCAGCAGACAGTGTGATACGTGTGTCAGTGTGAGC 364  
|||||  
Cp 185 TTCAACATCAGGCGCTTATCATCAGCAGACAGTGTGATACGTGTGTCAGTGTGAGC 126  
|||||  
Db 365 AGCTGTTTCGGATTCTGGGCAAGACCAAGAAAGATGGGAATGCCAAGATACAGAGAC 424  
|||||  
Cp 125 AGCTGTTTCGGATTCTGGGCAAGACCAAGAAAGATGGGAATGCCAAGATACAGAGAC 66  
|||||  
Db 425 TGCTAAGAACTTAAATGTCG 443  
|||||  
Cp 65 TGCTAAGAACTTAAATGTCG 47
```

```
RESULT 6  
LOCUS A1139456 402 bp mRNA EST 23-SEP-1998  
DEFINITION gc20901.x1 Soares fetal_heart_NBH19W Homo sapiens cDNA clone  
IMAGE:1710192 3' similar to contains element MER40 repetitive  
element; , mRNA sequence.  
A1139456  
g3645428
```

```
ACCESSION A1139456  
NID g3645428  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 402)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)
```

```
JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: -40ml3 fwd. ET from Amerisham  
High quality sequence stop: 398.  
Location/Qualifiers  
1..402
```

FEATURES

source

```
/organism="Homo sapiens"  
/note="Organ: heart; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st  
strand cDNA was primed with a Not I - Oligo(dt) primer [5'  
TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3']/  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by  
M.Fatima Bonaldo. This library was constructed from the  
same fetus as the fetal lung library, Soares fetal lung  
NBH19W."  
/db_xref="taxon:9606"  
/clone="IMAGE:1710192"
```

```
/clone_lid="Soares_fetal_heart_NBH19W"  
/sex="unknown"  
/dev_stage="19 weeks"  
/lab_host="DH10B (ampicillin resistant)"  
BASE COUNT      115 a      72 c      115 g      100 t  
ORIGIN
```

```
Query Match      75.3%; Score 232; DB 18; Length 402;  
Best Local Similarity 98.3%; Pred. No. 0.00e+00;  
Matches 236; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Db 163 ACCATTGGGAGATCCCAACCCATTGGGTAAACCTGAATGCTTTACGACGAGTGT 222  
|||||  
Cp 308 ACCATTGGGAGATCCCAACCCATTGGGTAAACCTGAATGCTTTACGACGAGTGT 249  
|||||  
Db 223 AGAAGCAGCGGTGTTGCAAGTGTAGAGACAGACAGTGTGTCAGTGTGTCAGCAGT 282  
|||||  
Cp 248 AGAAGCAGCGGTGTTGCAAGTGTAGAGACAGACAGTGTGTCAGTGTGTCAGCAGT 189  
|||||  
Db 283 GATTTCACATCAGGCGCTTATCATCAGCAGACAGTGTGATACGTGTGTCAGTGTG 342  
|||||  
Cp 188 GATTTCACATCAGGCGCTTATCATCAGCAGACAGTGTGATACGTGTGTCAGTGTG 129  
|||||  
Db 343 AGCAGCTGTTGTCGATTCTGGGCAAGACCAAGTAAATGGAATGCCAAGATACAG 402  
|||||  
Cp 128 AGCAGCTGTTGTCGATTCTGGGCAAGACCAAGTAAATGGAATGCCAAGATACAG 69
```

```
Search completed: Sun Mar 7 03:23:33 1999  
Job time : 728 secs.
```


Mon Mar 8 15:41:59 1999

US-08-962-094-3.rge

ArThuy, L
08/06/2004
Seq. IDs 1-20

MPERCH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPERCH_un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 03:27:49 1999; Maspar time 347.59 Seconds
1339.364 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-3
Description: (1-197) from US08962094.seq
Perfect Score: 197
N.A. Sequence: 1 GTTTTACCCCAATGGCTGG.....AAATACTATGACACAT 197
Comp: CAAATGGGTTTACCCCAAC.....TTTATTGATATCTGTTGTA

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 602357 segs, 1181590623 bases x 2

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: emb157
1:em_ba 2:em_fun 3:em_htg 4:em_hun1 5:em_hun2 6:em_in
7:em_om 8:em_ov 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_un 15:em_vt
Database: genbank110
16:gb_ba1 17:gb_ba2 18:gb_htg 19:gb_in 20:gb_om 21:gb_ov
22:gb_pat 23:gb_ph 24:gb_pl1 25:gb_pl2 26:gb_pl3
27:gb_pr1 28:gb_pr2 29:gb_ro 30:gb_st 31:gb_sts 32:gb_sy
33:gb_un 34:gb_vt

Statistics: Mean 9.496; Variance 5.472; scale 1.735

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match Length	DB ID	Description	Pred. No.
------------	-----------	-------	--------------	-------	-------------	-----------

No matches found.
Search completed: Sun Mar 7 03:38:19 1999
Job time : 630 secs.

24

34

```

Mpsrch.n n.a. - n.a. database search, using Smith-Waterman algorithm
Sun Mar 7 03:48:19 1999: Maspar time 38.11 Seconds
Run on: 703.291 Million cell updates/sec

```

Title:	>US-08-962-094-3
Description:	(1-197) from US08962094.seq
Perfect Score:	197
N.A. Sequence:	1 GTTTACCCCAATGGTTGG.....AAATACTATGACACAT 197
Comp:	CAAAATGGTTTACCAACC.....TTTATTGATCTGTTGTA

```
Nmatch      STD :  Dbase 0;  Query 0
```

post-processing:	Minimum Match 50%
	Listing first 1000 summaries

```
Database:
n:genses32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40
```

Statistics: Mean 7.570; Variance 4.834; scale 1.566

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match Score	Length	DB	ID	Description	Pred. No.

No matches found.

Search completed: Sun Mar 7 03:51:06 1999
Job time : 167 secs.

24

25

(TM)

[illegible]

Title:	Description:	Perfect Score:	N.A. Sequence:	Comp:
>US-08-962-094-3 (1-197) from US08962094.seq	197	1	GTTTTACCCAAATGGGTTGG.....AAATAACTATGACACAT	197
			CAAAATGGGTTTACCCAAAC.....TTTATTGATTACTCGTTGTA	

```

, Nmatch      STD : Dbase 0; Query 0

```

Searched: 2275026 seqs, 895388244 bases x 4

Post-processing: Minimum Match 508
Listing first 1000 summaries

```
Database:      emb1-est56
               1:em_gss1 2:em_gss2 4:em_gss3
Database:      gerhark/est109
               5:gb_est1 6:gb_est10 7:gb_est11 8:gb_est12 9:gb_est13
               10:gb_est14 11:gb_est15 12:gb_est16 13:gb_est17
               14:gb_est18 15:gb_est19 16:gb_est20 17:gb_est21
               18:gb_est22 19:gb_est23 20:gb_est24 21:gb_est25 22:gb_est26
               23:gb_est27 24:gb_est28 25:gb_est29 26:gb_gss1 27:gb_gss2
               28:gb_gss3 29:gb_gss4
```

Statistics: Mean 9.792; Variance 2.447; scale 4.001

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	196	99.5	402	18	A1139456	qc20g91.x1 Soares_feta	3.81e-265
2	195	99.0	422	21	W72838	zds7912.s1 Soares_feta	1.71e-226
3	168	85.3	436	21	W72837	zds7912.r1 Soares_feta	4.86e-226

Note: Post-processor removed 997 summaries from list due to search parameters chosen

ALIGNMENTS

RESULT	1	EST	23-SEP-1998
LOCUS	A1139456	402 bp	mRNA
DEFINITION	gc20901.x1 Soares, fetal heart, NBDH19W Homo sapiens beta -casein		
	IMAGE:1710192 3'	similar to contains element MERA0	repetitive

element // mRNA sequence

ACCESSION	A1139456
NID	g3645428
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Eumetazoa;

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 402)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
Tumor Gene Index
Unpublished (1997)

COMMENT

COMMENT

Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through NLM; contact the
NIH/NIH Consortium (info@image.llnl.gov) for further information.
seq. primer: -40n13 fwd. ET from Amersham
high quality sequence stop: 398.
location/Qualifiers

```

FEATURES
source
1. .402
/organism="Homo sapiens"
/organism="Homo sapiens" (Pharmacia) with

```

```

source
"/organism:"Homo sapiens"
"/organism:"Organ heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-GGGAGCGAGCCGCCATCTTTTTCCTTTTTTTT-3'] TTAACCATCTGAACGTGGAGCGGCCGCATCTTTTTCCTTTTTTTT-3') double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites through one round of normalization to a Cot = 5 library constructed by M.Patima Bonaldo. This library was constructed from the whole fetus as the fetal lung library, Soares fetal lung NBHL19W."
/db_xref="taxon:9606"
"/clone_id="IMAGE:1710192"
"/clone_lib="Soares_fetal_heart_NBHL19W"
"/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
lab 72 c 115 g 100 t

```

Query Match	99.5%	Score 196	DB 16	Accession
Best Local Similarity 100.0%		Pred. No. 3.1le-269		
Matches 196; Conservative		0; Mismatches	0; Indels	0; Gaps
Db	3	TGTTGCTCATAGTATTATTTTATTTGAAAGAAATTAACGTATTAGATTAAGGGGATA	62	
Cp	196	TGTTGCTCATAGTATTATTTTATTTGAAAGAAATTAACGTATTAGATTAAGGGGATA	13	
Db	63	TGCTTGGCGCAGAGTAAGTATGTGATGATAATCAGAGAAAGCATAAATAGTTGACCAATT	12	
Cp	136	TGCTTGGCGCAGAGTAAGTATGTGATGATAATCAGAGAAAGCATAAATAGTTGACCAATT	77	
Db	123	GCAGAAAGACTAACACTATTTCATCTCAGAGGACACACTCTACACTTTGGGAGATCCCCAA	18	
Cp	76	GCAGAAAGACTAACACTATTTCATCTCAGAGGACACACTCTACACTTTGGGAGATCCCCAA	17	
Db	183	CCCATTTGGGTAAAC	198	
Cp	16	CCCATTTGGGTAAAC	1	

RESULT	2	433 bp	mRNA	EST	16-OCT-1996
LOCUS	W72838				
DEFINITION	M557612.1 S1 Soares fetal heart NBH119W Homo sapiens cDNA clone				
	344865.3, similar to contains element MENT40 repetitive element ;				
	mRNA sequence.				
ACCESSION	W72838				
NTID	91382814				

KEYWORDS

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

EST.
human.
Homo sapiens
Eukaryotes: mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 432)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE

JOURNAL

COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 508 Std Error: 0.00
Seq primer: mob.RBGA+ET
High quality sequence stop: 394.
Location/Qualifiers
1. 432

FEATURES

SOURCE

/organism="Homo sapiens"
/note="Organ: heart; Vector: p773D (Pharmacia) with a
modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGAGGAGGAGCGCCGCGCATCTTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M. Fatima Bernaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."
/db_xref="taxon:9606"
/clone_id="344806"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
complement(<1..>432)
BASE COUNT 128 a 78 c 124 g 102 t
ORIGIN

Query Match 99.0%; Score 195; DB 21; Length 432;
Best Local Similarity 100.0%; Pred. No. 1,71e-267;
Matches 195; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 GTTGCTCATAGTATTTTATTTGAAAGAAATAAAGTATGAGATTAAGGGGATAT 60
|||||
Cp 195 GTTGCTCATAGTATTTTATTTGAAAGAAATAAAGTATGAGATTAAGGGGATAT 136
|||||
Db 61 CGTAGCAGAGGTAAGTATGATGATGAATACAGAGAGCATGAATAGTGTGACAAATG 120
|||||
Cp 135 CGTAGCAGAGGTAAGTATGATGATGAATACAGAGAGCATGAATAGTGTGACAAATG 120
|||||
Db 121 CAGAAGACTCAGAGTATCATCTCAGGACACACTTACATTTGGGAGATCCCAAC 180
|||||
Cp 75 CAGAAGACTCAGAGTATCATCTCAGGACACACTTACATTTGGGAGATCCCAAC 180
|||||
Db 181 CCATTGGGTAAC 195
|||||
Cp 15 CCATTGGGTAAC 1

RESULT 3

LOCUS W72837 436 bp mRNA EST 16-OCT-1996
DEFINITION zds7g12.1 Soares fetal heart NBHL19W Homo sapiens cDNA clone
344806 5' similar to contigs element-MR40=replicative element ;
RNA sequence.
ACCESSION W72837
NID g182813
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

AUTHORS

Eukaryotes: mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 436)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE

JOURNAL

COMMENT

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert length: 508 Std Error: 0.00
Seq primer: mob.RBGA+ET.
Location/Qualifiers
1. 436

FEATURES

SOURCE

/organism="Homo sapiens"
/note="Organ: heart; Vector: p773D (Pharmacia) with a
modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAGAGGAGGAGCGCCGCGCATCTTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M. Fatima Bernaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHL19W."
/db_xref="taxon:9606"
/clone_id="344806"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
complement(<1..>436)
BASE COUNT 92 a 128 c 81 g 134 t 1 others
ORIGIN

Query Match 85.3%; Score 168; DB 21; Length 436;
Best Local Similarity 99.4%; Pred. No. 4.86e-223;
Matches 174; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Db 262 GTTTTACCAATGGGTTGGGATCTCCGAATGATGAGATGTCCTGAGATGAATC 321
|||||
Cp 1 GTTTTACCAATGGGTTGGGATCTCCGAATGATGAGATGTCCTGAGATGAATC 60
|||||
Db 322 AGCTTGAGCTTCTGCAATNGGTCACAACTATTCATGCTTCCTGATTCATCACT 381
|||||
Cp 61 AGCTTGAGCTTCTGCAATNGGTCACAACTATTCATGCTTCCTGATTCATCACT 381
|||||
Db 382 ACTTACCTGCTTACGATATCCCTTATCTCTATTCAGTTATTTCTTTCAAA 436
|||||
Cp 120 ACTTACCTGCTTACGATATCCCTTATCTCTATTCAGTTATTTCTTTCAAA 174

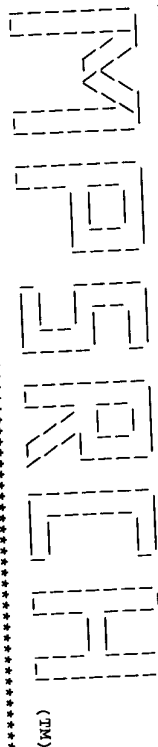
Mon Mar 8 15:42:01 1999

Search completed: Sun Mar 7 03:47:54 1999
Job time : 552 secs.

US-08-962-094-3.rst

Mon Mar 8 15:42:02 1999

US-08-962-094-4.rge



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_un n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sun Mar 7 03:51:30 1999; Maspar time 778.00 Seconds
1464.084 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-4
Description: (1-482) from US08962094.seq
Perfect score: 482
N.A. Sequence: 1 CGGCTCGAGCTCTTAGCTT.....AAATTAATGATGACACAT 482
Comp: GCCGAGCTCGAATCGAA.....TTTATGATACTCTGTGA

Scoring table: TABLE default
Gap 6

. Nmatch STD : Dbase 0; Query 0

Searched: 602357 seqs, 1181590623 bases x 2

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: emb157
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_com 8:em_ov 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_un 15:em_vl
Database: genbank110
16:gb_bal 17:gb_ba2 18:gb_htg 19:gb_in 20:gb_com 21:gb_ov
22:gb_pat 23:gb_ph 24:gb_pl1 25:gb_pl2 26:gb_pl3
27:gb_pr1 28:gb_pr2 29:gb_ro 30:gb_st 31:gb_sts 32:gb_sy
33:gb_un 34:gb_vl

Statistics: Mean 10.452; Variance 5.845; scale 1.788
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description	Pred. No.

No matches found.
Search completed: Sun Mar 7 04:09:11 1999
JOB time : 1061 secs.

Mon Mar 8 15:42:03 1999

US-08-962-094-4.rng

Page 1

MIRCH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_n n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Sun Mar 7 04:24:51 1999; MasPar time 86.95 Seconds
754.228 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-4
Description: (1-482) from US08962094.seq
Perfect Score: 482
N.A. Sequence: 1 CGGCTCGAGCTCTTACGCTT.....AAATACATGAGCAACAT 482
Comp: GCCGAGCTCGAGATCCGAA.....TTTATTGATCTCTTGA

Scoring table: TABLE default
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: n-geneseg32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 8.488; Variance 5.519; scale 1.538

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result Query
No. Score Match Length DB ID Description Pred. No.

No matches found.

Search completed: Sun Mar 7 04:28:26 1999

Job time : 215 secs.

7

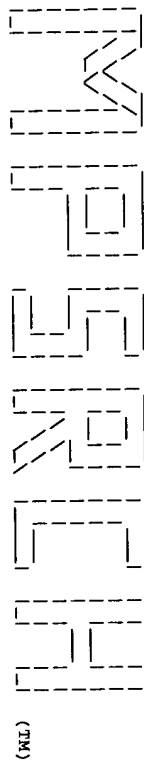
7

Mon Mar 8 15:42:04 1999

US-08-962-094-4.rst

Cp 143 GGAGCAGCTGTTGTGGAATTCGGGAGAGAGACGAAAGATGGAACCTCCCAAGATTACC 84
Db 420 AGGACTGCTAAGAACTTATGCTG 443
|||||
Cp 83 AGGACTGCTAAGAACTTATGCTG 60

Search completed: Sun Mar 7 04:24:28 1999
Job time : 891 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mperch_un n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 04:28:50 1999; Maspar time 883.30 Seconds
1479.490 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-5
Description: (1-553) from US08962094.seq
Perfect Score: 553
N.A. Sequence: 1 GAATTGGGCTCGAGCGGCTC.....AAAAAARGGCGCGCGC 553
Comp: CTTAACCCTCGAGCGCGAG.....TTTTTTTCCCGCGCGCG

Scoring table: TABLE default
Gap 6

Nmatch SMD : Dbase 0; Query 0

Searched: 602357 seqs, 1181590623 bases x 2

Post-processing: Minimum Match 508

Listing first 1000 summaries

Database: embi57
1:em_ba 2:em_fun 3:em_htg 4:em_hum1 5:em_hum2 6:em_in
7:em_om 8:em_ov 9:em_ov 10:em_pat 11:em_ph 12:em_pl
13:em_ro 14:em_un 15:em_v1
Database: genbank110
16:gb_ba1 17:gb_ba2 18:gb_htg 19:gb_in 20:gb_om 21:gb_ov
22:gb_pat 23:gb_ph 24:gb_pl1 25:gb_pl2 26:gb_pr1
27:gb_pr2 28:gb_pr3 29:gb_ro 30:gb_st 31:gb_sts 32:gb_sy
33:gb_un 34:gb_v1

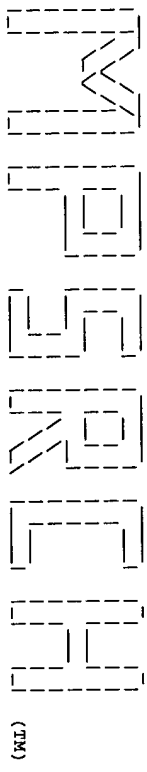
Statistics: Mean 11.381; Variance 19.039; scale 0.598

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
No matches found.						

Search completed: Sun Mar 7 04:48:00 1999
Job time : 1150 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_n n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:04:17 1999; Maspar time 86.58 Seconds
868.945 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-5
Description: (1-553) from US08962094.seq
Perfect Score: 553
N.A. Sequence: 1 GAATTCGGCTCGAGCGGCTC.....AAAAAAAAARGGCGCGCCGC 553
Comp: CTTAGCCGAGCTCGCCGAG.....TTTCTTTTTCGCCGCGCGCG

Scoring table: TABLE default
Gap 6

Mmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 68026449 bases x 2

Post-processing: Minimum Match 508
Listing first 1000 summaries

Database: n-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39 40:part40

Statistics: Mean 8.751; Variance 8.509; scale 1.028

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description	Pred. No.
-----	-----	-----	-----	-----	-----	-----

No matches found.

Search completed: Sun Mar 7 05:07:52 1999
Job time : 215 secs.


```

OY 229 GCTGTCGACACACATCGCAGCCATCTGCTCTCTACCATCTGCACACACCGCTGCTTACCC 278
DB 241 ACTGCTCGTAAAGACATTCACATTTTACCACCAATGGTGGGAGATCTCCGATGGTGA 300
OY 279 ACTGCTCGTAAAGACATTCACATTTTACCACCAATGGTGGGAGATCTCCGATGGTGA 338
DB 301 GGTGTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
OY 339 GGTGTCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 397
DB 361 TTCCGTGATTCATCACAATCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
OY 398 TTCCGTGATTCATCACAATCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 457
DB 421 TTATTTCTTTCCAAA 436
OY 458 TTATTTCTTTCCAAA 473

```

```

RESULT 2 W72838 432 bp mRNA EST 16-OCT-1996
LOCUS zdf7912.s1 Soares fetal heart NBHH19W Homo sapiens cDNA clone
DEFINITION 344806 3' similar to contains element MERA40 repetitive element ;,
mRNA sequence.
ACCESSION W72838
NID G1382814
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Eukaryotes; Metazoa; Chordata;
Vertebrata; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 432)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hulman, M., Kucaba, T., Le, M., Lennon, G., Matra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P., and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

```

```

TITLE Contact: Wilson RK
JOURNAL WashU-Merck EST Project
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 508 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 394.
Location/Qualifiers

```

```

FEATURES
source
1..432

```

```

/organism="Homo sapiens"
/Note="Organ: heart; Vector: p773D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TGTTACCATCTGAGTGGAGCGCGCCGATCTTTTCTTTTCTTTT 3')
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773D vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M. Fatima Bernaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHH19W.
/db_xref="taxon:9606"
/clone="344806"
/clone.lib="Soares fetal heart NBHH19W"
/sex="unknown"
/dev_stage="19 weeks"

```

```

mRNA /lab_host="DH10B (ampicillin resistant)"
BASE COUNT 128 a 78 c 124 g 102 t
complement(<1..>432)
ORIGIN

```

```

Query Match 76.98; Score 425; DB 21; Length 432;
Best Local Similarity 99.8%; Pred. No. 1,14e-221;
Matches 431; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

```

DB 1 GTTGTCTATGATTTATTTTATTTTAAAGAAATTAATGATAGATTAAGGGGATAT 60
CP 494 GTTGTCTATGATTTATTTTATTTTAAAGAAATTAATGATAGATTAAGGGGATAT 435
DB 61 CGTAGCAAGGTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 120
CP 434 CGTAGCAAGGTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 375
DB 121 CAGAGACTCAAGCTGATTCATCTCAGGAGACACCTCTACCATTCGGAGATCCCAAC 180
CP 374 CAGAGACTCAAGCTGATTCATCTCAGGAGACACCTCTACCATTCGGAGATCCCAAC 315
DB 181 CCATTTGGGTAAAGTGAATGTCTTTACGAGCAGTGTGTAGACAGCGGTGTCAGT 240
CP 314 CCATTTGGGTAAAGTGAATGTCTTTACGAGCAGTGTGTAGACAGCGGTGTCAGT 255
DB 241 GGTAGAGACAGCAGTGTGTCGAGTGTGTCAGCAGCAGTGTGTCAGCAGCAGTGTGTC 300
CP 254 GGTAGAGACAGCAGTGTGTCGAGTGTGTCAGCAGCAGTGTGTCAGCAGCAGTGTGTC 195
DB 301 ATCATCAGAGGACCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 360
CP 194 ATCATCAGAGGACCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 136
DB 361 GGGCAGAGACCGAAGATGGAATCTCCAGAGTACAGAGCTGTAGAACTTCATGG 420
CP 136 GGGCAGAGACCGAAGATGGAATCTCCAGAGTACAGAGCTGTAGAACTTCATGG 76
DB 421 TGTGACATGAA 432
CP 75 TGTGACATGAA 64

```

```

RESULT 3 A1139456 402 bp mRNA EST 23-SEP-1998
LOCUS qc20901.x1 Soares fetal heart NBHH19W Homo sapiens cDNA clone
DEFINITION IMAGE:1710192 3' similar to contains element MERA40 repetitive
element ;, mRNA sequence.
ACCESSION A1139456
NID G3645428
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Euthera;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 402)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT

```

```

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 398.
Location/Qualifiers
1..402

```

```

FEATURES
source
/organism="Homo sapiens"
/Note="Organ: heart; Vector: p773D (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'

```


TTTACCAATCTAGTGGAGCGGCCCATCTTTTTTTTTTTTTTTT 3',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia) digested with Not I and cloned into
the Not I and Eco RI sites of a modified pTZ19 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M. Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBH19W."
/db_xref="taxon:9606"
/clone="IMAGE:1710192"
/clone_lib="Soares_fetal_heart_NBH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"

BASE COUNT 115 a 72 c 115 g 100 t
ORIGIN

Query Match 71.6%; Score 396; DB 18; Length 402;
Best Local Similarity 99.3%; Pred. No. 1,35e-204;
Matches 399; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 1 TTTGTCCTCATGTTATTTTATTTGAAAGAAATAACTGATTAGATAAGAGGA 60
|||||
Cp 497 TTGTTGCTCAGTATTTATTTTATTTGAAAGAAATAACTGATTAGATAAGAGGA 438
|||||

Db 61 TATCTAGGCAAGTATGATTTGATGTAATCAGAGAGCATGATGTTGACCA 120
|||||
Cp 437 TATCTAGGCAAGTATGATTTGATGTAATCAGAGAGCATGATGTTGACCA 378
|||||

Db 121 TTGCAAGAGACTCAAGTATTCATCTCAGGAGACACTTACCATTTGGAGATCCCC 180
|||||
Cp 377 TTGCAAGAGACTCAAGTATTCATCTCAGGAGACACTTACCATTTGGAGATCCCC 318
|||||

Db 181 AACCCATTGGGTAACATGATGCTTTTACGACAGTGTGTAGAGCAGCGGTGTC 240
|||||
Cp 317 AACCCATTGGGTAACATGATGCTTTTACGACAGTGTGTAGAGCAGCGGTGTC 258
|||||

Db 241 AGTGGTTGAGCAGCAGTGTGCGAGTGTTCAGCAGCAGTGTGTTTCAGCATCAGGGGC 300
|||||
Cp 257 AGTGGTTGAGCAGCAGTGTGCGAGTGTTCAGCAGCAGTGTGTTTCAGCATCAGGGGC 198
|||||

Db 301 TTTCATTACGAGGAGCAGCAGTGTGATGCTGTGATGATGAGCAGCAGTGTGATG 360
|||||
Cp 197 TTTCATTACGAGGAGCAGCAGTGTGATGCTGTGATGATGAGCAGCAGTGTGATG 138
|||||

Db 361 CTGGGAGAGACAGATGATGGAAGTCCCAAGAGTACAG 402
|||||
Cp 137 CTGGGAGAGACAGATGATGGAAGTCCCAAGAGTACAG 96
|||||

RESULT 4
LOCUS R75793 403 bp mRNA EST 06-JUN-1995
DEFINITION Y121f09.r1 Homo sapiens cDNA clone 158921 5' similar to contains
LTR3 repetitive element ;.
ACCESSION R75793
NID 9850475
KEYWORDS EST.

ORGANISM
Homo sapiens
Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 403)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, N., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, B., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE The Mashu-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

Contact: Wilson RK
Mashu-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 307
Source: IMAGE Consortium, LNL.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (Info@image.lnl.gov) for further information.
Location/Qualifiers

FEATURES
source 1..403
/organism="Homo sapiens"
/clone="158921"

BASE COUNT 82 a 113 c 89 g 116 t 3 others
ORIGIN

Query Match 57.1%; Score 316; DB 16; Length 403;
Best Local Similarity 96.2%; Pred. No. 1.10e-157;
Matches 379; Conservative 0; Mismatches 5; Indels 10; Gaps 10;

Db 2 GCATTTTGTCTGTCTGCTCCCTATCTTCAAGTCACACCATGAAATTTTACAGTCTG 61
|||||
Cp 39 GCATTTTGTCTGTCTGCTCCCTATCTTCAAGTCACACCATGAAATTTTACAGTCTG 98
|||||

Db 62 GNACTCTGGGAGTTCATCTTCTGCTGCTGCGCAAGATCGCAACAGCTGCTCA 121
|||||
Cp 99 GTACTCTTGGAGTTTCATCTTCTGCTGCTGCGCAAGATCGCAACAGCTGCTCA 158
|||||

Db 122 GCTGACAGTATCAGCTACTGCTGCTGCTGATGATGAAGCCCTGATGCTGAACACT 181
|||||
Cp 159 GCTGACAGTATCAGCTACTGCTGCTGCTGATGATGAAGCCCTGATGCTGAACACT 218
|||||

Db 182 GCTGCTGCAACCTGCGACACACTGCTGCTGCTTACACTGCAACACCGCTGCTTACC 241
|||||
Cp 219 GCTGCTGCAACCTGCGACACACTGCTGCTGCTTACACTGCAACACCGCTGCTTACC 278
|||||

Db 242 ACTGCTCGTAAAGATTCCTCAATTTACCCAAATGGGTTGGGATCTCCGAATGGTAG 301
|||||
Cp 279 ACTGCTCGTAAAGATTCCTCAATTTACCCAAATGGGTTGGGATCTCCGAATGG-TAG 337
|||||

Db 302 GAGTGTGCTCCGAGAGATGGGAATCAGCTTGATGCTTCTGGCAATGGGTACCAANTAT 361
|||||
Cp 338 -AGTGTGCTCCGAG-ATGG-ATTCAGCTTGATGCTT-CTG-CAATTTGG-TCACAACTAT 391
|||||

Db 362 TTCAATGCTTCTGATGATTTTCAATTAAT 395
|||||
Cp 392 T-CATGCTTCTGCTG-ATT-CATCAACTACTT 422
|||||

RESULT 5
LOCUS AA340069 294 bp mRNA EST 21-APR-1997
DEFINITION EST45219 Fetal skin Homo sapiens cDNA 5' end similar to similar to
spot 2, submandibular gland protein, mRNA sequence.
ACCESSION AA340069
NID 91992307
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotes; Eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae;
Homo.
REFERENCE 1 (bases 1 to 294)

Mon Mar 8 15:42:06 1999

US-08-962-094-5.rst

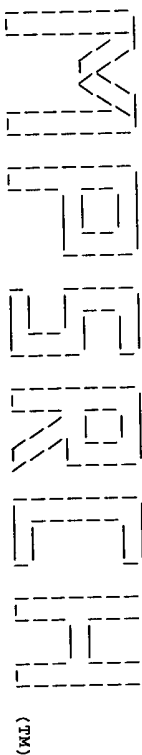
Page 5

CP 157 GGAGCAGCTGTGTGGATTCTGGGACAGACCAAGAAAGATGGAACCTCCACAGGTACC 98

DB 420 AGGACTGCTAAGACTTAATGCTG 443

CP 97 AGGACTGCTAAGACTTCATGCTG 74

Search completed: Sun Mar 7 05:03:54 1999
Job time : 932 secs.



 Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Msearch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:12:24 1999; Maspar time 9.19 Seconds
 68.651 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-17
 Description: (1-39) from US08962094.pep
 Perfect Score: 213
 Sequence: 1 AONPTTAAPADTYPATGPADDEAPDAETAAATTATATTA 39

Scoring table: PAM 150
 Gap 11

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 50%
 Listing first 1000 summaries

Database: a-geneseq32
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29

Statistics: Mean 19.949; Variance 100.669; scale 0.198

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description	Pred. No.

No matches found.

Search completed: Sun Mar 7 05:14:27 1999
 Job time : 123 secs.

W P E R L H (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MSPCH_PP protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:11:00 1999; MasPar time 4.13 Seconds
353.996 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-17
Description: (1-39) from US08962094.pep
Perfect Score: 213
Sequence: 1 AQNPTTAAPADTYPATGPDADDEAPDAETTAATTATTAA 39
Scoring table: PAM 150
Gap 11

Searched: 116738 seqs, 37463448 residues

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: p1r58
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 28.767; Variance 75.764; scale 0.380
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
---------------	-------	----------------	--------	----	-------------	-----------

No matches found.

Search completed: Sun Mar 7 05:12:02 1999
Job time : 62 secs.

 M P E R I H
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_gp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:08:14 1999; Maspar time 3.87 Seconds
 270.715 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-17
 Description: (1-39) from US08962094.pep
 Perfect Score: 213
 Sequence: 1 AQNPTTAAPADTYPATGPADDEAPDAETTAATTATTAA 39

Scoring table: PAM 150
 Gap 11

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 50%
 Listing first 1000 summaries

Database: swiss-prot35
 1:swissprot

Statistics: Mean 29.901; Variance 73.335; scale 0.408

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Query Length	ID	Description	Pred. No.
No matches found.					

Search completed: Sun Mar 7 05:09:15 1999
 Job time : 61 secs.

WIPSENI (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:09:35 1999; MasPar time 8.40 Seconds

Tabular output not generated. 231.126 Million cell updates/sec

Title: >US-08-962-094-17
Description: (1-39) from US08962094.pep
Perfect Score: 213
Sequence: 1 AQNPTAAPADTYPATGPADDEAPDAETTAAATTAATAA 39

Scoring table: PAM 150
Gap 11

Searched: 165420 seqs, 49795644 residues

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: sptrembl6
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 28.755; Variance 72.941; scale 0.394

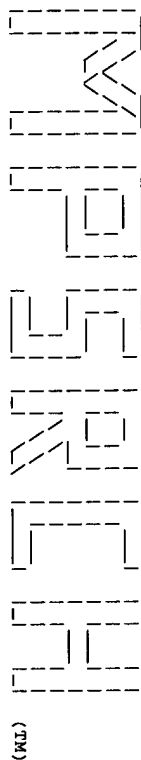
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
---------------	-------	----------------	--------	----	-------------	-----------

No matches found.

Search completed: Sun Mar 7 05:10:41 1999
Job time : 66 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:20:12 1999; Maspar time 4.17 Seconds
151.362 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-18
Description: (1-39) from US08962094.pep
Perfect Score: 252
Sequence: 1 TTATTAAPTTATTAATASTARKDIPVLPKRWGDLPNRGVC 39

Scoring table: PAM 150
Gap 11

Searched: 131922 seqs, 16180660 residues

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: a-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29

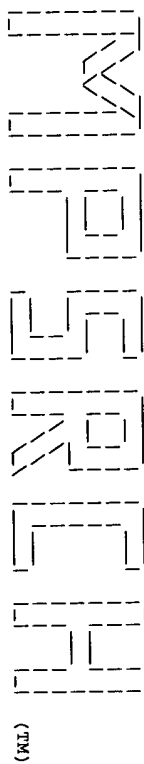
Statistics: Mean 23.465; Variance 92.489; scale 0.254
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	ID	Description	Pred. No.
No.	Score	Match	Length	DB

No matches found.

Search completed: Sun Mar 7 05:22:07 1999
Job time : 115 secs.



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPerch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:18:28 1999; Maspar time 5.95 Seconds
245.440 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-18
Description: (1-39) from US08962094.pep
Perfect Score: 252
Sequence: 1 TTATTAAPTTTATTASTTARKDIPVLPKWGDLPNGRVC 39

Scoring table: PAM 150
Gap 11

Searched: 116738 seqs, 37463448 residues

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: p1r58
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 32.408; Variance 66.743; scale 0.486

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
No matches found.						

Search completed: Sun Mar 7 05:19:49 1999
Job time : 81 secs.

WIPSEH (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:14:47 1999; MasPar time 2.82 Seconds
370.807 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-18
Description: (1-39) from US08962094.pep
Perfect Score: 252
Sequence: 1 TTATTAAPTTATTASTARKDIPVLPKWGDLPNGRYC 39

Scoring table: PAM 150
Gap 11

Searched: 74019 segs, 26840295 residues

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: swiss-prot35
1:swissprot

Statistics: Mean 33.471; Variance 60.737; scale 0.551

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.

No matches found.

Search completed: Sun Mar 7 05:16:17 1999
Job time : 90 secs.

10

Mon Mar 8 15:41:48 1999

US-08-962-094-18.rspt

Page 1

MPESETH
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpserch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:16:40 1999; Maspar time 9.93 Seconds
195.650 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-18
Description: (1-39) from US08962094.pep
Perfect Score: 252
Sequence: 1 TTATTAAPTTATTAASTARKDIPVLPKWGDLPNGRYC 39

Scoring table: PAM 150
Gap 11

Searched: 165420 segs, 49795644 residues

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: sptrembl6
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 32.963; Variance 61.313; scale 0.538

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Match	Length	ID	Description	Pred. No.
------------	-------	-------	--------	----	-------------	-----------

No matches found.

Search completed: Sun Mar 7 05:18:06 1999
Job time : 86 secs.

 M P S E Q I H
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_LP protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:26:24 1999; Maspar time 8.97 Seconds
 37.882 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-19
 Description: (1-21) from US08962094.pep
 Perfect Score: 166
 Sequence: 1 ARKDIPVLPRKVGDLPRNGRC 21

Scoring table: PAM 150
 Gap 15

Searched: 131922 segs, 16180660 residues

Post-processing: Minimum Match 50%
 Listing first 1000 summaries

Database: a-geneseq32
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
 14:part14 15:part15 16:part16 17:part17 18:part18
 19:part19 20:part20 21:part21 22:part22 23:part23
 24:part24 25:part25 26:part26 27:part27 28:part28
 29:part29

Statistics: Mean 21.655; Variance 69.163; scale 0.313

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length DB	ID	Description	Pred. No.

No matches found.

Search completed: Sun Mar 7 05:28:03 1999
 Job time : 99 secs.

WATERMAN (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msearch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:25:03 1999; Maspar time 4.79 Seconds
164.193 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-19
Description: (1-21) from US08962094.pep
Perfect Score: 166
Sequence: 1 ARKDIPVLPKMWGDLPMGRVC 21

Scoring table: PAM 150
Gap 15

Searched: 116738 segs, 37463448 residues

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: p1r58
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 29.205; Variance 46.987; scale 0.622

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description	Pred. No.
---------------	----------------	--------	----	-------------	-----------

No matches found.

Search completed: Sun Mar 7 05:26:01 1999
Job time : 58 secs.

Mon Mar 8 15:41:51 1999

US-08-962-094-19.rsp

Page 1

M P S R L H
(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:22:30 1999; Maspar time 3.54 Seconds
159.000 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-19
Description: (1-21) from US08962094.pep
Perfect Score: 166
Sequence: 1 ARKDIPVLPKMWGDLPNGRVC 21

Scoring table: PAM 150
Gap 15

Searched: 74019 seqs, 26840295 residues

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: swiss-prot35
1:swissprot

Statistics: Mean 30.068; Variance 42.251; scale 0.712

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

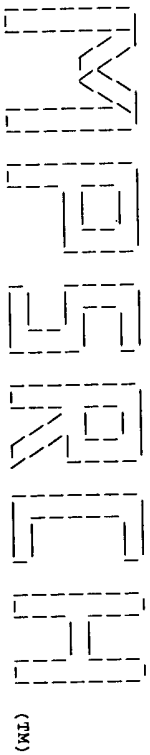
SUMMARIES

Result No.	Query Match	Length	ID	Description	Pred. No.
------------	-------------	--------	----	-------------	-----------

No matches found.

Search completed: Sun Mar 7 05:23:23 1999
job time : 53 secs.

4



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mparch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:23:45 1999; Maspas time 8.49 Seconds
123.224 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-19
Description: (1-21) from US08962094.pap
Perfect Score: 166
Sequence: 1 ARKDIPVLPKWKVGDLPNGRVC 21

Scoring table: PAM 150
Gap 15

Searched: 165420 seqs, 49795644 residues

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: sptrembl6
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phage 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 29.368; Variance 42.447; scale 0.692
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description	Pred. No.

No matches found.

Search completed: Sun Mar 7 05:24:42 1999
Job time : 57 secs.

10



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Sun Mar 7 05:32:48 1999; MasPar time 8.81 Seconds
38.578 Million cell updates/sec

Tabular output not generated.

Title: >US-08-962-094-20
Description: (1-21) from US08962094.dep
Perfect Score: 125
Sequence: 1 AAPADTYPATGPADDEAPDAE 21

Scoring table: PAM 150
Gap 15

Searched: 131922 segs, 16180660 residues

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: a-geneseq32
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29

Statistics: Mean 18.207; Variance 65.743; scale 0.277

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

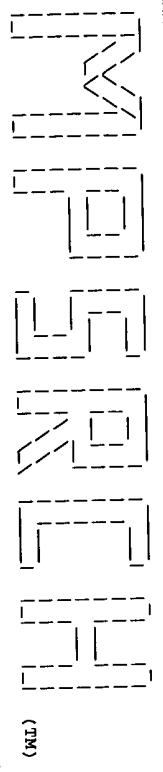
Result	Query	Score	Match	Length	ID	Description	Pred. No.

No matches found.

Search completed: Sun Mar 7 05:34:39 1999
Job time : 111 secs.

3

3



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch.p protein - protein database search, using Smith-Waterman algorithm
Run on: Sun Mar 7 05:31:12 1999; Maspar time 4.87 Seconds
161.680 Million cell updates/sec
Tabular output not generated.

Title: >US-08-962-094-20
Description: (1-21) from US08962094.pep
Perfect Score: 125
Sequence: 1 AAPADTYPATGPADEAPDAE 21

Scoring table: PAM 150
Gap 15

Searched: 116738 seqs, 37463448 residues
Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: p1r58
1: p1r1 2: p1r2 3: p1r3 4: p1r4
Statistics: Mean 25.550; Variance 41.666; scale 0.613

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description	Pred. No.
1	63	50.4	825	2 JCA163 DNA-binding protein 5	1.10e+00

Note: Post-processor removed 999 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ENTRY JCA163 #type complete
TITLE DNA-binding protein 5B5 - rat
ORGANISM #formal_name Rattus norvegicus #common_name Norway rat
DATE 27-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 07-Nov-1997
ACCESSION JCA163; PC4040
REFERENCE JCA163
#authors Suzuki, E.; Kojima, N.; Yoshimura, K.; Uyemura, K.; Odate, K.; Akagawa, K.
#journal J. Biochem. (1995) 118:122-128
#title Cloning and sequence analysis of cDNA for a possible DNA-binding protein 5B5 in the nervous system.
#accession JCA163
##molecule_type mRNA
##residues 1-825 #label SUZ

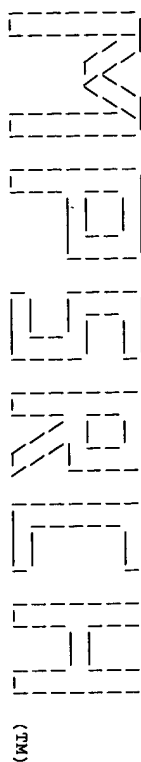
##cross-references DBJ:D37934; NID:9531260; PID:dl007730; PID:9531261
##experimental_source brain
#accession PC4040
##molecule_type protein
#residues 230-455 #label SUZ
COMMENT This protein has an abundance of arginine, a glycine-rich region and a proline cluster. This protein has a DNA-binding ability and is expressed especially in neurons.
KEYWORDS nerve; phosphoprotein
FEATURE
436-443 #region nuclear location signal
722-731 #region proline cluster
62,258,345,360,404,
413,570,613,635,
752,820
#binding_site phosphate (Ser) (covalent) (by protein kinase C) #status predicted
#length 825 #molecular_weight 86831 #checksum 7778

SUMMARY
Query Match 50.4%; Score 63; DB 2; Length 825;
Best Local Similarity 37.5%; Pred. No. 1.10e+00;
Matches 6; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

Db 138 SPEGAFFGCGPAREEG 153
QY 2 AAPADTYPATGPADEA 17

Search completed: Sun Mar 7 05:32:26 1999
Job time : 74 secs.

11



Release 3.1a John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Mserch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Sun Mar 7 05:28:24 1999; Mspartime 3.92 Seconds
143.779 Million cell updates/sec
Tabular output not generated.

Title: >US-08-962-094-20
Description: (1-21) from US08962094.pep
Perfect Score: 125
Sequence: 1 AAPADTYPATGPADDEAPDAE 21

Scoring table: PAM 150
Gap 15

Searched: 74019 segs, 26840295 residues

Post-processing: Minimum Match 50%
Listing first 1000 summaries

Database: swiss-prot35
1:swissprot

Statistics: Mean 26.391; Variance 38.782; scale 0.681

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	66	52.8	690	1 RHO_MICLU	TRANSCRIPTION TERMINAT	1.21e-01
2	63	50.4	825	1 5E5_RAT	5E5 ANTIGEN.	4.36e-01

Note: Post-processor removed 998 summaries from list due to search parameters chosen.

ALIGNMENTS

RESULT 1
ID RHO_MICLU STANDARD; PRT; 690 AA.
AC P52154;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE TRANSCRIPTION TERMINATION FACTOR RHO.
GN RHO.
OS MICROCOCCUS LUTEUS (MICROCOCCUS LYODEIKTICUS).
OC PROKARYOTA; FIRMICUTES; COCCI; MICROCOCCACEAE.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-5 AND 289-297.
RC STRAIN-EM;
RX MEDLINE; 96132802.
RA NOMATZKE W.L., RICHARDSON J.P.;

RL J. BIOL. CHEM. 271:742-747(1996).
RN [2]
RP SEQUENCE OF 205-690 FROM N.A.
RC STRAIN-EM;
RX MEDLINE; 94327472.
RA OPERMAN T., RICHARDSON J.P.;
RL J. BACTERIOL. 176:5033-5043(1994).
RN [3]
RP REVISION TO 500.
RC STRAIN-EM;
RA NOMATZKE W.L.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: FACILITATES TRANSCRIPTION TERMINATION BY A MECHANISM
CC THAT INVOLVES RHO BINDING TO THE NASCENT RNA, ACTIVATION OF RHO'S
CC RNA-DEPENDENT ATPASE ACTIVITY, AND RELEASE OF THE RNA FROM THE
CC DNA TEMPLATE. RNA-DEPENDENT NTASE WHICH UTILIZES ALL FOUR
CC RIBONUCLEOSIDE TRIPHOSPHATES AS WELL AS DATP AS SUBSTRATES, BUT
CC HAS A SIGNIFICANT LOWER ACTIVITY WITH CTP.
CC -1- SUBUNIT: HOMOHETEROMER (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
CC EMBL: L27277; G166540;
DR TRANSCRIPTION TERMINATION; HELICASE; ATP-BINDING; RNA-BINDING.
KW INIT_MET 0
FT DOMAIN 30 35 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 318 321 RNA-BINDING (RNP1) (BY SIMILARITY).
FT NP_BIND 441 448 ATP (POTENTIAL).
FT CONFLICT 291 291 G -> P (IN AA SEQUENCE).
SQ SEQUENCE 690 AA; 75030 MM; 93AB0634 CRC32;

Query Match 52.8%; Score 66; DB 1; Length 690;
Best Local Similarity 47.6%; Pred. No. 1.21e-01;
Matches 10; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

DB 71 AAPAETAPAAASSEDAPAAE 91
QY 1 AAPADTYPATGPADDEAPDAE 21

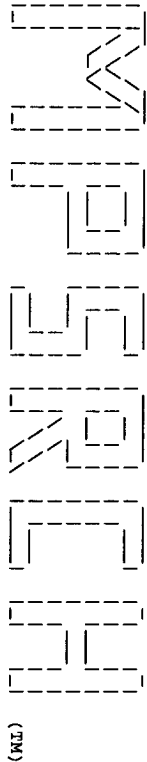
RESULT 2
ID 5E5_RAT STANDARD; PRT; 825 AA.
AC Q63003;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE 5E5 ANTIGEN.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WISTAR; TISSUE-BRAIN;
RX MEDLINE; 96015159.
RA SUZUKI E., KOJIMA N., YOSHIMURA K., UYEMURA K., OBATA K., AKAGAWA K.;
RL J. BIOCHEM. 118:122-128(1995).
CC -1- FUNCTION: MIGHT HAVE DNA-BINDING ABILITY.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEURONS.
DR EMBL; D37934; G531261;
KW DNA-BINDING; NUCLEAR PROTEIN; ANTIGEN.
SQ SEQUENCE 825 AA; 86831 MM; BC01D1EF CRC32;

Query Match 50.4%; Score 63; DB 1; Length 825;
Best Local Similarity 37.5%; Pred. No. 4.36e-01;
Matches 6; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

DB 138 SPEGAPGGGPAEEG 153
QY 2 AAPADTYPATGPADDEA 17

Search completed: Sun Mar 7 05:29:13 1999
Job time : 49 secs.

10



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msearch_p protein - protein database search, using Smith-Waterman algorithm
Run on: Sun Mar 7 05:29:33 1999; MasPar time 5.15 Seconds
Tabular output not generated.
Title: >US-08-962-094-20
Description: (1-21) from US08962094.1.ppt
Perfect Score: 125
Sequence: 1 AAPADTYPATGPADDEAPDAE 21
Scoring table: PAM 150
Gap 15
Searched: 165420 segs, 49795644 residues
Post-processing: Minimum Match 508
Listing first 1000 summaries
Database: sprenb16
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
13:sp.vertebrate 14:sp.virus
Statistics: Mean 25.336; Variance 37.686; scale 0.672
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES
Result No. Score Match Length DB ID Description Pred. No.
1 64 51.2 1463 14 069298 MAJOR IMMEDIATE EARLY 3.62e-01
2 63 50.4 801 5 023635 SIMILAR TO LONG TANDEM 5.56e-01
3 63 50.4 3158 2 052819 PCZA363.3. 5.56e-01
Note: Post-processor removed 997 summaries from list due to search parameters chosen.
ALIGNMENTS
RESULT 1
ID 069298 PRELIMINARY; PRT; 1463 AA.
AC 069298;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
DE MAJOR IMMEDIATE EARLY PROTEIN.
GN
GN
OS
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; GAMMAHERPESVIRIDAE.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-SA-2;
RX MEDLINE: 95282504.
RA JOHNSON M.A., TYACK S.G., PRIDEAUX C., KONGSUTAN K., SHEPPARD M.;
RL VIRUS RES. 35:193-204(1995).
DR EMBL: L32139; G493598; -;
SQ SEQUENCE 1463 AA; 158701 MW; CA15F280 CRC32;
Query Match 51.2%; Score 64; DB 14; Length 1463;
Best Local Similarity 64.3%; Pred. No. 3.62e-01;
Matches 9; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
Db 947 SGPAGSYATGPAD 960
QY 1 AAPADTYPATGPAD 14
RESULT 2
ID 023635 PRELIMINARY; PRT; 801 AA.
AC 023635;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE SIMILAR TO LONG TANDEM REPEAT REGION OF SLALIDASE.
GN ZK84.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; ACCELLOMATES; NEMATODA; SECERNENTEA; RHABDITIDA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA BURTON J., CONNELL M., COPSET T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DORBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKIN L., ROOPE A., SANDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULLIVAN J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA KIRSTEN J.;
RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: U23181; G726436; -;
SQ SEQUENCE 801 AA; 77123 MW; 7288A89D CRC32;
Query Match 50.4%; Score 63; DB 5; Length 801;
Best Local Similarity 42.9%; Pred. No. 5.56e-01;
Matches 9; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
Db 411 SAAAEETPAPAPAEETPAPE 431
QY 1 AAPADTYPATGPADDEAPDAE 21
RESULT 3
ID 052819 PRELIMINARY; PRT; 3158 AA.
AC 052819;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
DE PCZA363.3.
OS AMYCOLAPOPSIS ORIENTALIS.
OC EUBACTERIA; FIRMICUTES; ACTINOMYCETES; PSEUDONOCARDIACEAE;
OC AMYCOLAPOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RA VAN WAGENINGEN A., KIRKPATRICK P., WILLIAMS D., HARRIS B., KERSHAW J.,
RA LENNARD N., JONES M., JONES S., SOLENERG P.;

RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.

DR EMBL AJ223999; E1231243; -

DR PROSITE: PS00455; AMP_BINDING; 3.

SQ SEQUENCE 3158 AA; 336315 MW; E162473A CRC32;

Query Match 50.4%; Score 63; DB 2; Length 3158;

Best Local Similarity 47.1%; Pred. NO. 5.56e-01; Indels 0; Gaps 0;

Matches 8; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

Db 2430 ENYTPAAGDEAADAQ 2446

OY 5 DTYPATCPADDEAPDAE 21

Search completed: Sun Mar 7 05:30:51 1999
Job time : 78 secs.